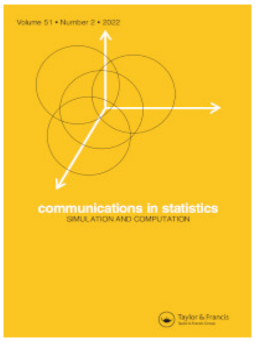


Taylor & Francis
Taylor & Francis Group

Communications in Statistics - Simulation and Computation

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/lssp20

The adaptive LASSO regression and empirical mode decomposition algorithm for enhancing modelling accuracy

Abdullah S. Al-Jawarneh & Mohd. Tahir Ismail

COMMUNICATIONS IN STATISTICS - SIMULATION AND COMPUTATION®
https://doi.org/10.1080/03610918.2022.2032154

The adaptive LASSO regression and empirical mode decomposition algorithm for enhancing modelling accuracy

Abdullah S. Al-Jawarneh[a] and Mohd. Tahir Ismail[b]

[a]Faculty of Science, Department of Mathematics, Jerash University, Jerash, Jordan; [b]School of Mathematical Sciences, Universiti Sains Malaysia, Pulau Pinang, Malaysia

ABSTRACT

The first part of the Hilbert–Huang transformation is named the empirical mode decomposition (EMD). Which employed to decompose the non-stationary and non-linear time series dataset into a finite set of orthogonal decomposition components. These components have been used in several studies as the new predictor variables to predict the behavior of the response variables. Adaptive LASSO (AdLASSO) regression is a technical penalized regression method used to determine the most relevant predictors on the response variable with achieving the consistency in terms of variable selection and ensuring that they are asymptotically normal. Hence, the main objective of this study is to apply the proposed EMD-AdLASSO method involving two cases of initial weights to identify the decomposed components that exhibit the strongest effects to produce a consistent model and to improve the prediction accuracy. The simulation study and real dataset used the daily exchange rate dataset of three countries against the US dollar are applied. The results showed that the proposed method in the two cases of the initial weight outperformed other existing methods by effectively identifying the decomposition components, with high prediction accuracy. This is primarily observed in the case of using the ridge regression method based on the EMD as the initial weight in the proposed EMD-AdLASSO method.

ARTICLE HISTORY
Received 23 January 2021
Accepted 17 January 2022

KEYWORDS
Adaptive LASSO; empirical mode decomposition; oracle property; selection variables; intrinsic mode function; time series

1. Introduction

Traditional decomposition methods assume that the time series dataset should be either stationary or linear before the analysis process. For example, Fourier decomposition (Titchmarsh 1948) and wavelet decomposition (Chan 1995) methods. However, most real-life data appear as non-stationary and nonlinear time series datasets. Thus, the decomposition of non-stationary and nonlinear time series are important issues that need to be considered in conducting an analysis. Meanwhile, there is a lack of analytical methodologies that employed to deal with these time-series datasets.

In 1998, Huang et al. became interested in the distorted non-stationary and nonlinear time-series signals without moving from the time domain into the frequency domain, and the information was maintained in the time domain. This led to a new technical method called the empirical mode decomposition (EMD) to be proposed (Huang 2014). The EMD method does not require a pre-condition for the time series dataset, unlike the traditional methods (Huang 2014). Currently, the EMD has been widely used in various fields of science, including medicine (Masselot et al.

CONTACT Abdullah S. Al-Jawarneh abdssj2004@yahoo.com; abdullah.jawarneh86@gmail.com Faculty of Science, Department of Mathematics, Jerash University, Jerash, Jordan.

2018), electronic engineering (Suvasini et al. 2015), civil and construction engineering (OBrien, Malekjafarian, and González 2017), economics (Jaber, Ismail, and Altaher 2014), and environmental science (Naik, Satapathy, and Dash 2018).

The practical principle of the EMD is it aims to separate the non-stationary and nonlinear signal into a finite set of orthogonal non-overlapping time scale components, namely, the intrinsic mode functions and residual components (decomposition components) (Moore et al. 2018). Each of the components is different in terms of its physical form (i.e., wavelength and frequency). It includes information particular frequencies found within the original time series dataset (Qin et al. 2016; Al-Jawarneh et al. 2020; Al-Jawarneh and Ismail 2021). As such, the decomposition components can be used as new predictor variables to predict their effects and behaviors about other response variables using suitable models. For instance, the ordinary least-squares (OLS) regression and forward stepwise regression methods are applied based on the EMD method by (Yang, Tsai, and Huang 2011), using combined stepwise regressions (SR) based on the EMD method (Adarsh and Janga Reddy 2019; Zhao et al. 2018), the least absolute shrinkage and selection operator (LASSO) regression based on the EMD method (Qin et al. 2016; Masselot et al. 2018), the ridge regression (RR) based on the decomposition components via EMD method by (Naik, Satapathy, and Dash 2018; Ali et al. 2019) and the Elastic-net regression based on the EMD method by (Al-Jawarneh et al. 2020; Al-Jawarneh, Ismail, and Awajan 2021; Al-Jawarneh and Ismail 2021).

The results obtained from the use of the SR method lacked reliability in selecting the optimal model (Smith 2018). In contrast, the OLS estimate has a low prediction accuracy and face difficulty in reducing the number of predictor variables. While the RR method (Hoerl and Kennard 1970) still cannot deal with the reduction of the predictor numbers, so that the unnecessary predictor variables will still exist in the final model. The LASSO method (Tibshirani 1996) is inconsistent for variable selection. That means the method does not have the oracle property (Fan and Li 2001; Zou 2006).

The purpose of this study aims to identify the decomposed components that exhibit the most substantial effects on the response variable and improve prediction accuracy. The unbiased method, known as the adaptive LASSO (AdLASSO) regression (Zou 2006) method based on the EMD algorithm is applied.

The organization of this article is as follows. Section 2 will present a description of the EMD method, adaptive LASSO regression, and the proposed EMD-AdLASSO method involving two cases for the initial adaptive weights. Section 3 will provide the simulation study and the daily exchange rate time series dataset applied in this study. Section 4 will illustrate the analysis and discussion of the results. Finally, a conclusion for the study is provided in Section 5.

2. Methodology

In this section, we discussed the used methods in detail. Firstly, the EMD method was selected to deal with non-stationary and nonlinear predictors via the sifting process algorithm. Second, the technical penalized least squares estimator known as the adaptive LASSO regression method will be applied. Finally, the proposed EMD-AdLASSO method will be presented.

2.1. Empirical mode decomposition

The empirical mode decomposition (EMD) proposed by Huang et al., in 1998, EMD is a new technique of the decompose method and represents the first part of the Hilbert–Huang transform (Huang et al. 1998). The EMD aims to break the original signal $x(t)$ which is a non-stationary and nonlinear time series dataset into a finite set of nearly orthogonal decomposition components, which keeps the time domain of the signal unchanged (Huang 2014). These decomposition

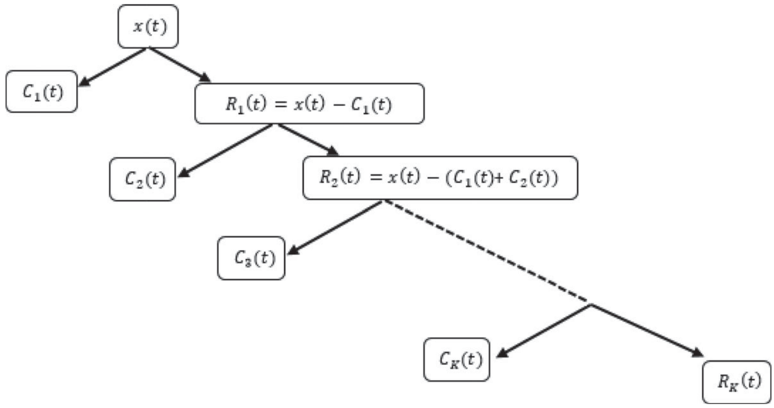


Figure 1. The tree graph of the Empirical mode decomposition.

components are called the intrinsic mode function (IMF) $\{C_k(t); k = 1, 2, ..., K\}$ components and one residual $R(t)$ component represents the trend of $x(t)$ (Niu and Wang 2014; Moore et al. 2018). The IMF's components should satisfy the following conditions: (a) The numbers of local extrema (i.e., maximum and minimum) and zero crossings must be equal or differ at most by one, (b) The mean envelope among the upper and lower envelopes is equal to a zero value (Park, Cho, and Oh 2013; Huang 2014).

The EMD decomposes $x(t)$ using an iterative process which is called the sifting process (SP) algorithm, which is described as the following algorithm (Huang 2014) and Figure 1.

Algorithm: SP Algorithm

Input: $x(t)$

Output: $\{C_k(t);\ k = \overline{1, K}\}$ and $R(t)$ components

i. Suppose that: $x(t) = R_0(t)$; and $q = 1$ and $k = 1$.
ii. Determine the local extrema points.
iii. By the cubic spline curve, identify the envelope;
   - Upper envelope $u_q(t)$ of the local maximum points.
   - Lower envelope $l_q(t)$ of the local minimum points.
iv. Calculate the mean envelope: $m_q(t) = (u_q(t) + l_q(t))/2$.
v. Find the IMF components: $h_q(t) = x(t) - m_q(t)$.
vi. Check, if the $h_q(t)$ satisfies the IMF conditions;
   - *Yes*: Then $h_q(t) = C_k(t)$, save the output $C_k(t)$, and go to (vii).
   - *No*: Let $q = q + 1$, repeat from (ii) until (vi)
vii. Calculate $R_k(t) = R_{k-1}(t) - C_k(t)$.
viii. Check, if the $R_k(t)$ is a monotonic function or satisfy the stoppage criterion of the standard deviation SD_h;

$$SD_h = \sum_{t=0}^{T} \frac{\left(h_{q-1}(t) - h_q(t)\right)^2}{{h_{q-1}}^2(t)};\ 0.2 \leq SD_h \leq 0.3$$

   - *Yes*: Save $R(t)$, and go to **End**.
   - *No*: Let $k = k + 1$, repeat from (ii) until (viii).

End.

The original signal expressed as the linear combination of the k IMF components and one residual $R(t)$ as Equation (Huang 2014):

$$x(t) = \sum_{k=1}^{K} C_k(t) + R(t) \tag{1}$$

2.2. Adaptive LASSO regression

The adaptive LASSO regression (AdLASSO) was proposed by Zou in 2006 to treat the oracle property gap in the LASSO estimator using the weighting factors to control the bias (Zou 2006). The AdLASSO method is a weighted version of the LASSO method (Dicker, Huang, and Lin 2013). The principle of the AdLASSO estimator aims to eliminate the bias in the LASSO estimator using the weights to apply a different amount of shrinkage on the different coefficients. This means the small coefficients are shrunk more severely, while the large coefficients are shrinking less than the small coefficients. The consistency for variable selection with probability tending to one and asymptotically normal estimates will be achieved (Huang, Ma, and Zhang 2008; Wang et al. 2020; Pietrosanu et al. 2021). The AdLASSO estimator form is as follows:

$$\hat{\boldsymbol{\beta}}^{AdLASSO} = \arg\min_{\boldsymbol{\beta}} \left\{ \frac{1}{2n} \|\boldsymbol{y} - \hat{\boldsymbol{y}}\|_2^2 + AP_{\lambda,\gamma}(\boldsymbol{\beta}) \right\}; \; AP_{\lambda,\gamma}(\boldsymbol{\beta}) = \lambda \sum_{j=1}^{p} \omega_j \left|\beta_j\right| \tag{2}$$

where $[\boldsymbol{y}]_{n\times 1}$ is the response vector variable and $\hat{\boldsymbol{y}} = \boldsymbol{X}\hat{\boldsymbol{\beta}}$; $[\hat{\boldsymbol{y}}]_{n\times 1}$ is the estimated vector model, and $AP_{\lambda,\gamma}(\boldsymbol{\beta})$ is the AdLASSO penalty function, $\lambda > 0$ is the tuning parameter, $\omega_j > 0$; $j = 1, 2, ..., p$ is the adaptive weights of the j^{th} predictor which are defined as $\omega_j = 1/\left|\hat{\beta}_j^{init}\right|^{\gamma}$, where $\hat{\beta}_j^{init}$ is an initial estimate of the regression coefficient β_j obtained by the RR estimator ω^{RR} or OLS estimator ω^{OLS}, and $\gamma > 0$ is a positive constant (Wang et al. 2020). The optimal value of (λ, γ) can be chosen by the $K-$ fold cross-validation (CV) at $\gamma = \{0.5, 1, 2\}$ (Zou 2006).

Suppose that $D = \{j : \beta_j \neq 0\}$ denotes the set of regression coefficients $\boldsymbol{\beta}$ which have an important index effect, $D_n = \left\{j : \hat{\beta}_j^{AdLASSO} \neq 0\right\}$ denotes the set of $\boldsymbol{\beta}$ that is estimated using the AdLASSO method, $\lambda_n/\sqrt{n} \to 0$, and $\lambda_n n^{\gamma-1} \to \infty$ as $n \to \infty$, where λ_n varies with n. The AdLASSO has an oracle property; that is, the estimator $\hat{\boldsymbol{\beta}}^{AdLASSO}$ must satisfy the following properties:

i. Consistency for variable selection with probability tending to one:

$$\lim_{n\to\infty} p(\; D_n = D) = 1$$

ii. Asymptotic normality

$$\sqrt{n}(\hat{\boldsymbol{\beta}}_D^{AdLASSO} - \boldsymbol{\beta}_D)d \to N(0, \Sigma)$$

that is the $\hat{\boldsymbol{\beta}}_D^{AdLASSO}$ is a consistent estimator for $\boldsymbol{\beta}_D$ with normal limit and covariance matrix $var(\beta_D) = \Sigma$ (Fan and Li 2001; Zou 2006).

Based on the numerical optimization algorithm method, which is called the coordinate descent (COD) with the given λ and γ values, the AdLASSO estimates can be solved. The COD sub-problem is used to optimize each predictor. That solves exactly one predictor $\boldsymbol{X}_j$ while the rest of predictors $\boldsymbol{X}_f$ except the j^{th} predictor will be fixed (Friedman, Hastie, and Tibshirani 2010). Then, the problem can be written as follows:

$$\hat{\boldsymbol{\beta}}^{AdLASSO} = \arg\min_{\boldsymbol{\beta}} \left\{ \frac{1}{2n} \|\boldsymbol{y} - \boldsymbol{X}_f \boldsymbol{\beta}_f - \boldsymbol{X}_j \beta_j\|_2^2 + \lambda \frac{|\beta_j|}{\left|\hat{\beta}_j^{init}\right|^{\gamma}} + \lambda \sum_{f \neq j}^{p} \frac{|\beta_f|}{\left|\hat{\beta}_f^{init}\right|^{\gamma}} \right\} \tag{3}$$

Suppose $\boldsymbol{y} - \boldsymbol{X}_f \beta_f$ is the partial residual$\boldsymbol{r}_f$, then Equation (3) becomes as follows:

$$\hat{\boldsymbol{\beta}}^{AdLASSO} = \arg\min_{\boldsymbol{\beta}} \left\{ \frac{1}{2n} \|\boldsymbol{r}_f - \boldsymbol{X}_j \beta_j\|_2^2 + \lambda \frac{|\beta_j|}{\left|\hat{\beta}_j^{init}\right|^{\gamma}} + \lambda \sum_{f \neq j}^{p} \frac{|\beta_f|}{\left|\hat{\beta}_f^{init}\right|^{\gamma}} \right\} \tag{4}$$

Suppose that $\hat{\beta}_j$ is a solution of β_j, then the partial derivative of the Equation (4) with respect to β_j and equate the derivative to zero, that is:

$$-n^{-1} \boldsymbol{X}_j^t (\boldsymbol{r}_f - \boldsymbol{X}_j \hat{\beta}_j) + sign\left(\hat{\beta}_j\right) \frac{\lambda}{\left|\hat{\beta}_j^{init}\right|^{\gamma}} = 0 \tag{5}$$

Hence, $n^{-1} \boldsymbol{X}_j^t \boldsymbol{X}_j = 1$ assuming that the predictor variables are standardized. The AdLASSO coefficients are estimated using the following equation:

$$\hat{\beta}_j = n^{-1} X_j^t \boldsymbol{r}_f - sign\left(\hat{\beta}_j\right) \frac{\lambda}{\left|\hat{\beta}_j^{init}\right|^{\gamma}} \tag{6}$$

By the soft-thresholding function, the AdLASSO estimator in the Equation (6) for each $j = 1, 2, ..., p$ can be rewritten in the following form:

$$\hat{\beta}_j = S(n^{-1} X_j^t \boldsymbol{r}_f, \lambda\omega_j) = \begin{cases} n^{-1} X_j^t \boldsymbol{r}_f + \lambda\omega_j & \text{if} \quad n^{-1} X_j^t \boldsymbol{r}_f < \lambda\omega_j \\ 0 & \text{if} \quad \left| n^{-1} X_j^t \boldsymbol{r}_f \right| \leq \lambda\omega_j \\ n^{-1} X_j^t \boldsymbol{r}_f - \lambda\omega_j & \text{if} \quad n^{-1} X_j^t \boldsymbol{r}_f > \lambda\omega_j \end{cases}, \tag{7}$$

where $S(n^{-1} X_j^t \boldsymbol{r}_f, \lambda\omega_j)$ represents the soft-thresholding function.

2.3. Proposed EMD-AdLASSO method

The proposed EMD-AdLASSO method was presented in two cases for the initial adaptive weights as the EMD-OLS and the EMD-RR estimators. Next, the proposed method in the two cases, namely EMD-AdLASSO.OLS and EMD-AdLASSO.RR, were applied for multiple original predictor variables $x_j(t)$; $j = 1, 2, ..., p$. The proposed methods were designed as follows:

1) EMD-AdLASSO.OLS method:

i. Via the EMD method, the original signals $x_j(t)$ is decomposed into a finite set of the IMF components and one residual component for each separate j, it can be written as follows.

$$x_j(t) = \sum_{k=1}^{K} C_{j.k}(t) + R_j(t); j = 1, 2, ..., p \tag{8}$$

ii. Using all the decomposition components of the $x_j(t)$ in the first step, the new predictor variables to predict the behavior of the response variable$y(t)$ is as follows (Masselot et al. 2018):

$$\begin{aligned} y(t) &= C_{1.1}(t)\beta_{1.1} + C_{1.2}(t)\beta_{1.2} + \ldots + R_1(t)\beta_{1.K+1} + \ldots \\ &\quad + C_{j.1}(t)\beta_{j.1} + C_{j.2}(t)\beta_{j.2} + \ldots + R_j(t)\beta_{j.K+1} + \varepsilon(t) \\ &= \sum_{j=1}^{p}\left(\sum_{k=1}^{K} C_{j.k}(t)\beta_{j.k}\right) + R_j(t)\beta_{j.K+1} + \varepsilon(t) \end{aligned} \tag{9}$$

iii. Computing the initial weights $\omega^{EMD-OLS}$ based on the initial estimators EMD-OLS at $\gamma = 1$:

$$\omega_{j.k}^{EMD-OLS} = \left|\hat{\beta}_{j.k}^{EMD-OLS}\right|^{-1} \tag{10}$$

iv. Select the optimal tuning parameter value λ_{opt} by using the 10-fold CV method.

$$\lambda_{opt} = argmin_s\{CV_{\lambda_s}\};\ CV_{\lambda_s} = \frac{1}{10}\sum_{k=1}^{10} MSE_{k,\lambda_s,\omega^{EMD-OLS}};\ s = \overline{1,S} \tag{11}$$

v. Apply the AdLASSO method:

$$\hat{\boldsymbol{\beta}}_{EMD}^{AdLASSO.OLS} = \underset{\beta}{argmin}\left\{\frac{1}{2n}\sum_{i=1}^{n}\left(y(t) - \sum_{j=1}^{p}\left(\sum_{k=1}^{K} C_{j.k}(t)\beta_{j.k}\right) - R_j(t)\beta_{j.K+1}\right)^2 + OP_\lambda(\boldsymbol{\beta})\right\}; \tag{12}$$

$$OP_\lambda(\boldsymbol{\beta}) = \lambda_{opt}\sum_{j=1}^{p}\left(\sum_{k=1}^{K+1}\omega_{j.k}^{EMD-OLS}\left|\beta_{j.k}\right|\right)$$

2) EMD-AdLASSO.RR method:
Noted that steps (i) and (ii) are the same as in the previous EMD-LASSO.OLS method.
iii. Computing the initial weights ω^{EMD-RR} based on the initial estimators EMD-RR at $\gamma = 1$:

$$\omega_{j.k}^{EMD-RR} = \left|\hat{\beta}_{j.k}^{EMD-RR}\right|^{-1} \tag{13}$$

iv. Select the optimal tuning parameter value λ_{opt} by using the 10-fold CV method.

$$\lambda_{opt} = argmin_s\{CV_{\lambda_s}\};\ CV_{\lambda_s} = \frac{1}{10}\sum_{k=1}^{10} MSE_{k,\lambda_s,\omega^{EMD-RR}};\ s = \overline{1,S} \tag{14}$$

v. Apply the AdLASSO method:

$$\begin{aligned} \hat{\boldsymbol{\beta}}_{EMD}^{AdLASSO.RR} &= \underset{\beta}{argmin}\left\{\frac{1}{2n}\sum_{i=1}^{n}\left(y(t) - \sum_{j=1}^{p}\left(\sum_{k=1}^{K} C_{j.k}(t)\beta_{j.k}\right) - R_j(t)\beta_{j.K+1}\right)^2 + RP_\lambda(\boldsymbol{\beta})\right\}; \\ RP_\lambda(\boldsymbol{\beta}) &= \lambda_{opt}\sum_{j=1}^{p}\left(\sum_{k=1}^{K+1}\omega_{j.k}^{EMD-RR}\left|\beta_{j.k}\right|\right) \end{aligned} \tag{15}$$

Finally, a comparison was made between the proposed methods with traditional methods, namely, EMD-OLS, EMD-SR, EMD-RR, and EMD-LASSO. The comparison made was in terms of the decomposed components selection and the use of the residual sum of squares (RSS), root mean square error (RMSE), mean absolute error (MAE) and mean absolute scaled error (MASE) as the criteria tests:

Table 1. Mean performance criteria.

Method	RSS	RMSE	MAE	MASE
EMD-OLS	77.06964	1.0126870	0.8623016	0.7951836
EMD-SR	74.60106	0.9964267	0.8525577	0.7860674
EMD-RR	73.86578	0.9919956	0.8625673	0.7950496
EMD-LASSO	73.08547	0.9866494	0.8561660	0.7891685
EMD-ALASSO.OLS	**72.71307**	**0.9840945**	**0.8517993**	**0.7851780**
EMD-ALASSO.RR	**72.65568**	**0.9836985**	**0.8513228**	**0.7847640**

*Bold line values are the proposed methods

$$\mathrm{RSS} = \sum_{i=1}^{n}(y_i\hat{}y_i)^2 \qquad \mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i\hat{y}_i|$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i\hat{y}_i)^2} \qquad \mathrm{MASE} = \frac{1}{n}\sum_{i=1}^{n}\left[|y_i\hat{y}_i| \; / \left(\frac{1}{n1}\sum_{i=2}^{n}|y_iy_{i1}|\right)\right]$$

3. Application

This section implemented the numerical simulation experiment and a real dataset application to show the capacity of the proposed EMD-AdLASSO method. The analyses are made by R software that uses an EMD package proposed by (Kim and Oh 2009) to decompose nonstationary and nonlinear variables and the Glmnet package given by (Friedman, Hastie, and Tibshirani 2010) for the AdLASSO estimators.

3.1. Simulation study

The sine function was used to assess the ability of the proposed methods. The datasets were generated for the non-stationary and nonlinear multivariate predictors, and the response variable with the sample size $n = 250$ and the time-domain is $0 \leq t \leq 9$, by adding the white noise error $\varepsilon \sim iid\ N(0, 1)$ for the predictors (Qin et al. 2016; Al-Jawarneh et al. 2020; Al-Jawarneh, Ismail, and Awajan 2021; Al-Jawarneh and Ismail 2021). $3,000$ replications of the sample size of 250 were modeled. The 10-fold CV was used to estimate the optimal tuning parameter λ value. The formulas of the function test as follows:

$$y(t) = 0.5t + \sin(\pi t) + \sin(2\pi t) + \sin(6\pi t)$$

$$x_1(t) = 0.8t + \sin(0.3\pi t) + \sin(2\pi t) + \sin(7\pi t) + \sin(9\pi t) + \varepsilon$$

$$x_2(t) = 0.4t + \sin(0.2\pi t) + \sin(6\pi t) + \sin(5\pi t) + \sin(12\pi t) + \varepsilon$$

3.2. Daily exchange rate

The daily close exchange rates from 27/03/2015 to 25/10/2019 of three countries against the US dollar (USD) were applied in this study. Those selected countries were Japan (JAP), China (CHN), and Taiwan (TAW). All the datasets were collected from the Wall Street Journal database (https://www.wsj.com/).

In this application, the daily exchange rates of the JAP and CHN represented the original predictor variables, whereas the daily exchange rates of TAW was the response variable. The datasets were split into training and test datasets, where the training dataset represented the period

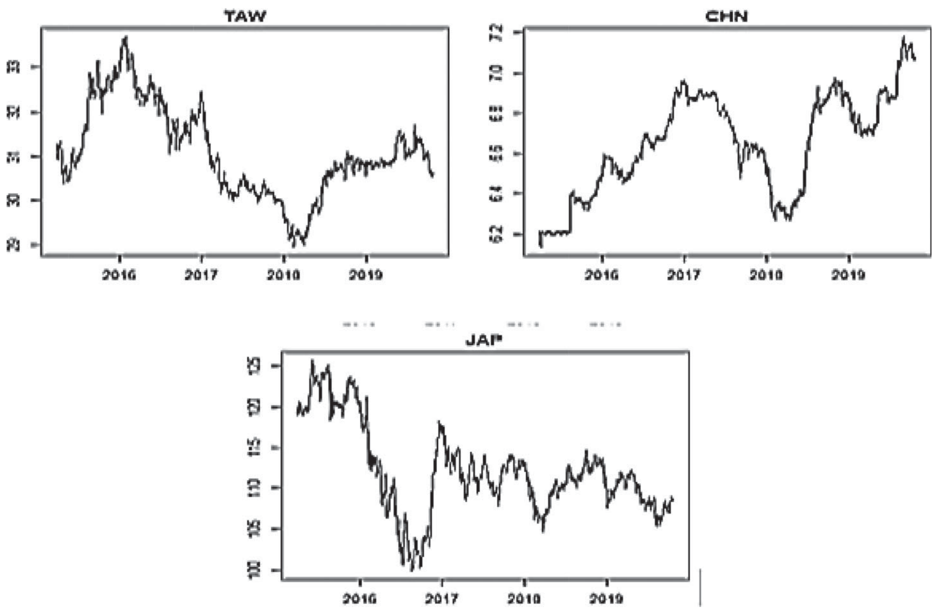


Figure 2. Plots of the original signals CHN and JAP and TAW.

between 27/03/2015 to 11/6/2018 to estimate the fit models. In contrast, the remaining datasets represented the test dataset for evaluating the performance criteria.

4. Results and discussion

The results of the analysis and discussion of the simulation experiment and actual time series data of the daily close exchange rates are presented in this section.

4.1. Simulation results

Table 1 describes the average of the performance criteria in terms of RSS, RMSE, MAE, and MASE for all the regression methods used in this study. The results show that the smallest error value in RSS, RMSE, MAE, and MASE is achieved using the EMD-AdLASSO.RR method and the second is achieved using the EMD-AdLASSO.OLS method. So, the proposed EMD-AdLASSO method in the two cases of the initial weight has achieved the highest level of reliability with high prediction accuracy in the simulation experiment.

4.2. Daily exchange rate results

Figure 2 illustrates the shape of the original primary predictors CHN, and JAP and the response TAW variables. The shape of the predictor variables and the response variable shows neither a constant value over time (i.e., the signal has a trend, changing levels, and seasonality rules that depend on the time) or fluctuates around the zero lines. Such conditions are called nonstationary and nonlinear conditions, respectively.

Figure 3 shows the decomposition components of the original predictors CHN and JAP via the EMD method. The first predictor CHN decomposes into seven IMFs $\{C_{1.1}, C_{1.2}, C_{1.3}, C_{1.4}, C_{1.5}, C_{1.6}, C_{1.7}\}$ and one residual R_1 components. The second predictor JAP has the same number of IMFs $\{C_{2.1}, C_{2.2}, C_{2.3}, C_{2.4}, C_{2.5}, C_{2.6}, C_{2.7}\}$ and one residual R_2

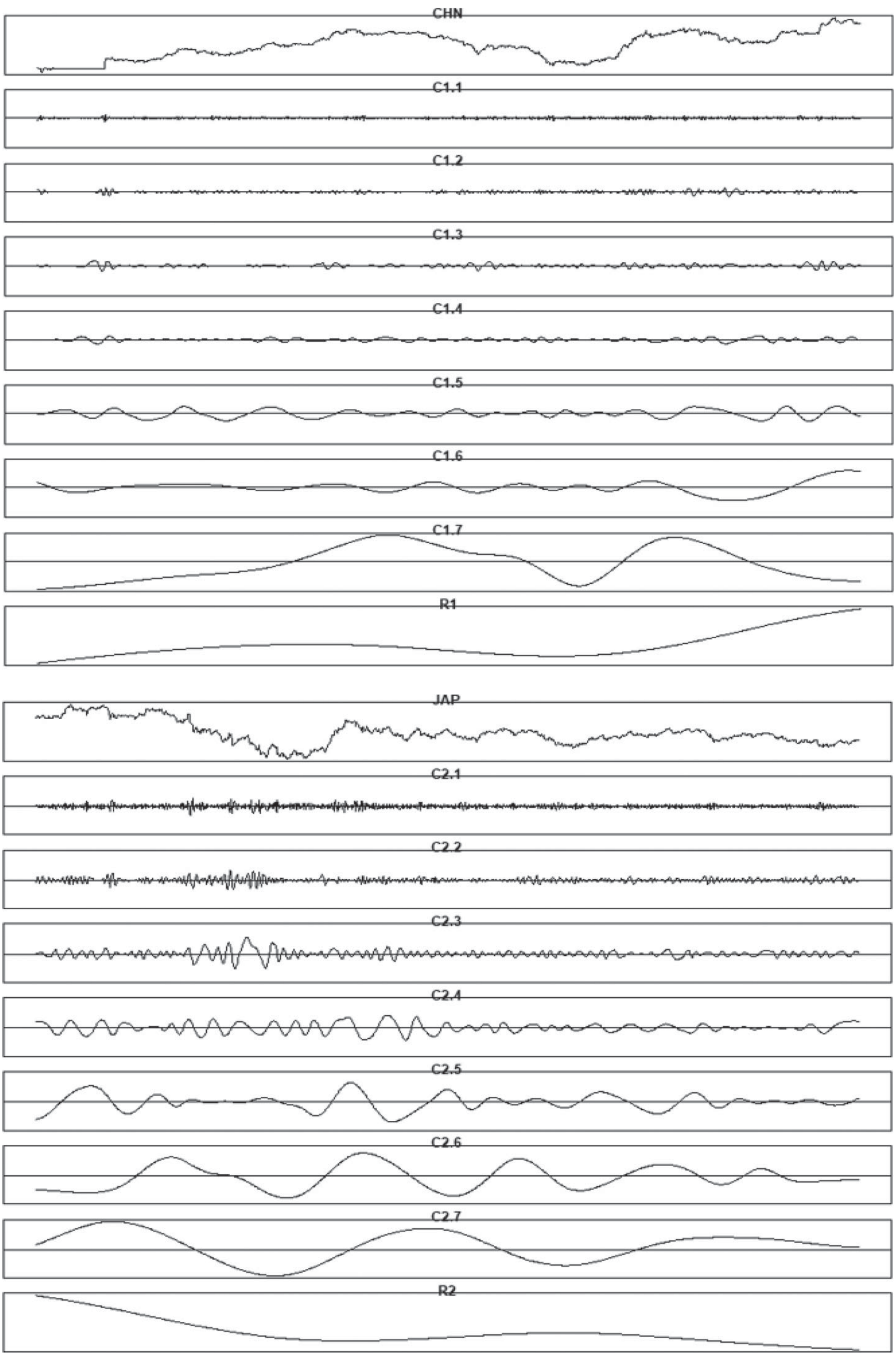
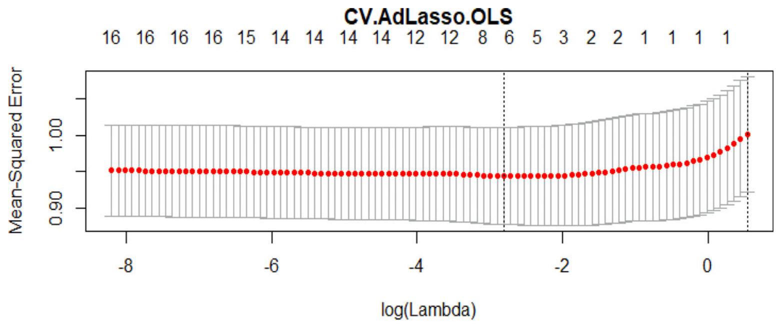


Figure 3. Decomposition of the original signals CHN and JAP via EMD.

components. Hence, each one of the decomposition components has differences in physical properties, particularly the wavelength and frequency compared with other decomposition components. For illustration, the first decomposition component in the two original predictors $C_{1.1}$ and



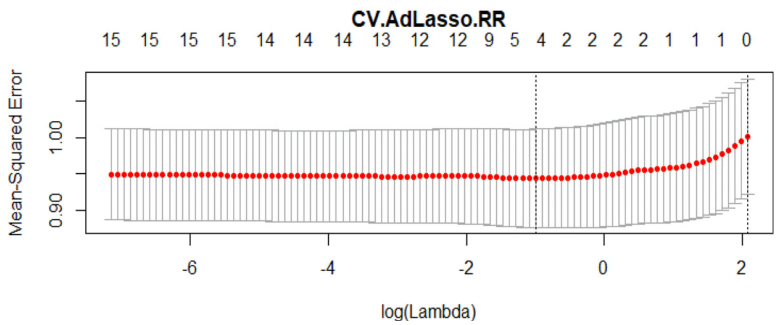


Figure 4. 10-fold CV estimation of the MSE as the $Log\ (\lambda)$ for the proposed method in two cases.

Table 2. Residual sum of squares error values.

Method	λ_{min}	Log(λ)	RSS
EMD-OLS			341.8256
EMD-SR			314.4687
EMD-RR	0.193699	−1.641449871	321.6946
EMD-LASSO	0.007293	−4.920889247	316.1166
EMD-AdLASSO.OLS	0.060464	−2.805719371	**313.1700**
EMD-AdLASSO.RR	0.369095	−0.996702028	**311.7273**

*Bold lines are the proposed methods and the smallest RSS values

Table 3. Coefficients estimation for the predictor variables.

	EMD-OLS	EMD-SR	EMD-RR	EMD-LASSO	EMD-AdLASSO.OLS	EMD-AdLASSO.RR
$\hat{\beta}_{1.1}$	−0.2391	−0.2289	−0.1974	−0.2298	−0.2234	−0.2208
$\hat{\beta}_{1.2}$	−0.1129	−0.1119	−0.0925	−0.1051	−0.0948	−0.0895
$\hat{\beta}_{1.3}$	−0.0419	0	−0.0300	−0.0333	0	0
$\hat{\beta}_{1.4}$	−0.0237	0	−0.0175	−0.0143	0	0
$\hat{\beta}_{1.5}$	−0.0347	0	−0.0293	−0.0290	0	0
$\hat{\beta}_{1.6}$	0.0080	0	0.0118	0.0056	0	0
$\hat{\beta}_{1.7}$	−0.1014	0	−0.0345	−0.0310	−0.0352	0
$\hat{\beta}_{1.8}$	0.0101	0	−0.0005	0	0	0
$\hat{\beta}_{2.1}$	−0.0600	−0.0574	−0.0545	−0.0538	−0.0276	−0.0209
$\hat{\beta}_{2.2}$	0.0361	0	0.0266	0.0267	0	0
$\hat{\beta}_{2.3}$	−0.0326	0	−0.0278	−0.0264	0	0
$\hat{\beta}_{2.4}$	0.0533	0	0.0394	0.0396	0	0
$\hat{\beta}_{2.5}$	0.0539	0	0.0458	0.0490	0.0109	0.0061
$\hat{\beta}_{2.6}$	−0.0391	0	−0.0336	−0.0317	0	0
$\hat{\beta}_{2.7}$	0.0856	−0.0555	0.0373	0.0345	0.0186	0
$\hat{\beta}_{2.8}$	−0.0756	0	−0.0048	0	0	0

*$\hat{\beta}_{j.k}$ is the *k* component coefficient of the original predictor *j*

Table 4. Performance criteria.

Method	λ_{min}	RMSE	MAE	MASE
EMD-OLS		0.9757871	0.7173476	0.6725151
EMD-SR		0.9359259	0.6764175	0.6341431
EMD-RR	0.193699	0.9466177	0.6853022	0.6424726
EMD-LASSO	0.007293	0.9383750	0.6796471	0.6371708
EMD-AdLASSO.OLS	0.060464	**0.9339913**	**0.6737997**	**0.6316889**
EMD-AdLASSO.RR	0.369095	**0.9318375**	**0.6717694**	**0.6297855**

*Bold lines are the proposed methods and the smallest values in term of performance criteria

$C_{2.1}$ has a short wavelength and highest frequency, while the $C_{1.7}$ and $C_{2.7}$ components have a long-wavelength and lower frequency.

Figure 4 shows the plots of the 10-fold CV of the AdLASSO.OLS, and AdLASSO.RR regression methods based on the EMD components for selecting the optimal λ. In each plot, the upper horizontal line represents the numbers of non-zero coefficients for a given λ. The vertical two dotted lines from left to right represent the optimal selected value of λ . The first vertical line is the location of λ value at the minimum mean squared error (MSE) called λ_{min}, and the second vertical line is the location of λ value at the MSE is within one standard error from the minimum is called λ_{1se}. Hence in Figure 4, the number of the non-zero coefficients will decrease as the value of the λ increases. For instance, in the AdLASSO.OLS method case, six non-zero coefficient components are chosen in the final mode at λ_{min}, while only one component at λ_{1se} will be chosen. However, in this study, the value of λ_{min} has been used to choose the non-zero coefficient components.

Table 2 displays the residual sum of squares error (RSS) values of the EMD-AdLASSO method in the two cases, compared with the traditional methods. The results of the RSS values show that the smallest value is achieved using EMD-AdLASSO.RR ($\lambda_{min} = 0.369095;\ \text{RSS} = 311.7273$). The second smallest RSS value is achieved using the EMD- AdLASSO.OLS ($\lambda_{min} = 0.060464;\ \text{RSS} = 313.17$). The RSS value provides the best methods to select and support the fitting regression models, where the proposed method in the two cases performs better with the smallest error than the other previous methods used.

Table 3 explains the estimation of the non-zero coefficient of the decomposition components that have the most effect on the response variable using the EMD-AdLASSO method in the two cases for the initial weights, and the current methods are similar to those used in previous studies. In the EMD-AdLASSO.OLS method, the number of the coefficients that have been selected equals to six non-zero coefficients which the $C_{1.1}, C_{1.2}, C_{1.7}, C_{2.1}, C_{2.5}$ and $C_{2.7}$, whereas in the EMD-AdLASSO.RR method four the non-zero coefficients selected are $C_{1.1}, C_{1.2}, C_{2.1}$, and $C_{2.5}$ and they have the most effect on the response variables. Hence, the components in the EMD-LASSO method, with the small coefficients, have larger weights and force to be equal to zero in the proposed methods (like $\hat{\beta}_{1.6}$ in the EMD-LASSO non-zero coefficient, while $\hat{\beta}_{1.6}$ equal zero in the proposed methods), while the large coefficients are shrinking less than the small coefficients depending on the initial weights.

Table 4 illustrates the performance criteria to measure the prediction accuracy of the proposed method. A comparison is also made between the current methods using the criteria tests, which are the RMSE, MAE, and MASE. The results show that EMD-AdLASSO.RR method provides the smallest error value in terms of RMSE, MAE, and MASE, and the second smallest error achieved by the EMD-AdLASSO.OLS method. The proposed EMD-AdLASSO method in the two cases provides better results than the traditional methods, especially in the EMD-AdLASSO.RR method.

5. Conclusions

In this study, we propose the adaptive LASSO regression based on the EMD method (EMD-AdLASSO) in two cases of the initial weight at EMD-OLS and EMD-RR that have been applied.

The novel proposes that a method is used to identify the relevance of the orthogonal decomposed components via EMD, which have the most effect on the response variable to produce a consistent model and to improve prediction accuracy.

The EMD algorithm has been used to decompose the non-stationary and non-linear original time series predictor into a finite set of the IMF components and one residual component. In contrast, the AdLASSO method, based on the OLS and RR as the initial weight, has been used for selecting the best decomposition components, which are the most significant in terms of their response variable.

The results of the simulation study and the close daily exchange rate dataset show that the proposed EMD-AdLASSO method in the two cases has efficiently chosen the actual decomposition components. These components have the strongest effects on the response variable with high prediction accuracy. Especially, in the EMD-RR case as the initial weight, where the small coefficient shrank more strictly. That leads to the reduction in the number of variables selection, and lower prediction error compared to the EMD-OLS case.

In this study, the EMD-AdLASSO method process has many advantages comparing to traditional methods: (1) The EMD algorithm makes the relationship between the variables more reliable in the sense of time and frequency domains simultaneously. (2) Each decomposition components extracted via EMD has information concerning a particular frequency that is found within the original predictor variable. (3) The AdLASSO regression effectively produced a consistent model by identifying the decomposition components that exhibited the strongest effects among the components. (4) The proposed EMD-AdLASSO method gives high prediction accuracy comparing to the traditional methods.

ORCID

Abdullah S. Al-Jawarneh http://orcid.org/0000-0001-7684-9017
Mohd. Tahir Ismail http://orcid.org/0000-0003-2747-054X

References

Adarsh, S., and M. Janga Reddy. 2019. Multiscale characterization and prediction of reservoir inflows using MEMD-SLR coupled approach. *Journal of Hydrologic Engineering* 24 (1):04018059. doi: 10.1061/(ASCE)HE.1943-5584.0001732.

Ali, M., R. C. Deo, T. Maraseni, and N. J. Downs. 2019. Improving SPI-derived drought forecasts incorporating synoptic-scale climate indices in multi-phase multivariate empirical mode decomposition model hybridized with simulated annealing and kernel ridge regression algorithms. *Journal of Hydrology* 576:164–84. doi: 10.1016/j.jhydrol.2019.06.032.

Al-Jawarneh, A. S., and M. T. Ismail. 2021. Elastic-net regression based on empirical mode decomposition for multivariate predictors. *Pertanika Journal of Science and Technology* 29 (1): 199–215. doi: 10.47836/pjst.29.1.11.

Al-Jawarneh, A. S., M. T. Ismail, and A. M. Awajan. 2021. Elastic net regression and empirical mode decomposition for enhancing accuracy of the model selection. *International Journal of Mathematical, Engineering and Management Sciences* 6 (2):564–83. doi: 10.33889/IJMEMS.2021.6.2.034.

Al-Jawarneh, A. S., M. T. Ismail, A. M. Awajan, and A. R. Alsayed. 2020. Improving accuracy models using elastic net regression approach based on empirical mode decomposition. *Communications in Statistics-Simulation and Computation*:1–20. doi: 10.1080/03610918.2020.1728319.

Chan, Y.T. 1995. *Wavelet basics*. New York, NY: Springer, Kluwer Academic Publishers.

Dicker, L., B. Huang, and X. Lin. 2013. Variable selection and estimation with the seamless-L 0 penalty. *Statistica Sinica* 23:929–62. doi: 10.5705/ss.2011.074.

Fan, J., and R. Li. 2001. Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American Statistical Association* 96 (456):1348–60. doi: 10.1198/016214501753382273.

Friedman, J., T. Hastie, and R. Tibshirani. 2010. Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software* 33 (1):1–22. doi: 10.18637/jss.v033.i01.

Hoerl, A. E., and R. W. Kennard. 1970. Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics* 12 (1):55–67. doi: 10.1080/00401706.1970.10488634.

Huang, N. E. 2014. Introduction to the Hilbert–Huang transform and its related mathematical problems. In *Hilbert–Huang transform and its applications*, 1–26, Singapore: World Scientific. doi: 10.1142/9789814508247_0001.

Huang, J., S. Ma, and C. H. Zhang. 2008. Adaptive Lasso for sparse high-dimensional regression models. *Statistica Sinica* 18:1603–18.

Huang, N. E., S. Zheng, R. L. Steven, M. C. Wu, H. H. Shih, Q. Zheng, N. C. Yen, C. C. Tung, and H. H. Liu. 1998. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 454 (1971): 903–95. doi: 10.1098/rspa.1998.0193.

Jaber, A. M., M. T. Ismail, and A. M. Altaher. 2014. Application of empirical mode decomposition with local linear quantile regression in financial time series forecasting. *The Scientific World Journal* 2014:708918doi.org/10.1155/2014/708918. doi: 10.1155/2014/708918.

Kim, D., and H. S. Oh. 2009. EMD: A package for empirical mode decomposition and Hilbert spectrum. *The R Journal* 1 (1):40–6. doi: 10.32614/RJ-2009-002.

Masselot, P., F. Chebana, D. Bélanger, A. St-Hilaire, B. Abdous, P. Gosselin, and T. B. Ouarda. 2018. EMD-regression for modelling multi-scale relationships, and application to weather-related cardiovascular mortality. *Science of the Total Environment* 612:1018–29. doi: 10.1016/j.scitotenv.2017.08.276.

Moore, K. J., M. Kurt, M. Eriten, D. M. McFarland, L. A. Bergman, and A. F. Vakakis. 2018. Wavelet-bounded empirical mode decomposition for measured time series analysis. *Mechanical Systems and Signal Processing* 99:14–29. doi: 10.1016/j.ymssp.2017.06.005.

Naik, J., P. Satapathy, and P. Dash. 2018. Short-term wind speed and wind power prediction using hybrid empirical mode decomposition and kernel ridge regression. *Applied Soft Computing* 70:1167–88. doi: 10.1016/j.asoc.2017.12.010.

Niu, H., and J. Wang. 2014. Phase and multifractality analyses of random price time series by finite-range interacting biased voter system. *Computational Statistics* 29 (5):1045–63. doi: 10.1007/s00180-014-0479-0.

OBrien, E. J., A. Malekjafarian, and A. González. 2017. Application of empirical mode decomposition to drive-by bridge damage detection. *European Journal of Mechanics - A/Solids* 61:151–63. doi: 10.1016/j.euromechsol.2016.09.009.

Park, M., S. Cho, and H. S. Oh. 2013. The role of functional data analysis for instantaneous frequency estimation. *Computational Statistics* 28 (5):1965–87. doi: 10.1007/s00180-012-0389-y.

Pietrosanu, M., J. Gao, L. Kong, B. Jiang, and D. Niu. 2021. Advanced algorithms for penalized quantile and composite quantile regression. *Computational Statistics* 36 :333–14. doi: 10.1007/s00180-020-01010-1.

Qin, L., S. Ma, J. C. Lin, and B. C. Shia. 2016. Lasso regression based on empirical mode decomposition. *Communications in Statistics - Simulation and Computation* 45 (4):1281–94. doi: 10.1080/03610918.2013.826361.

Smith, G. 2018. Step away from stepwise. *Journal of Big Data* 5 (1):32. doi: 10.1186/s40537-018-0143-6.

Suvasini, L., S. Prethivika, S. S. Murugan, and V. Natarajan. 2015. Extraction of binary sequences in a frequency shift keying-modulated signal by empirical mode decomposition algorithm against ambient noises in underwater acoustic channel. In *Artificial intelligence and evolutionary algorithms in engineering systems; Advances in Intelligent Systems and Computing*, 325: 371–8. Kumaracoil, India: Springer. doi: 10.1007/978-81-322-2135-7_40.

Tibshirani, R. 1996. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)* 58 (1):267–88. doi: 10.1111/j.2517-6161.1996.tb02080.x.

Titchmarsh, E. C. 1948. *Introduction to the theory of Fourier integrals*. 2nd ed. Oxford: Clarendon Press.

Wang, F., S. Mukherjee, S. Richardson, and S. M. Hill. 2020. High-dimensional regression in practice: An empirical study of finite-sample prediction, variable selection and ranking. *Statistics and Computing* 30 (3):697–719. doi: 10.1007/s11222-019-09914-9.

Yang, A. C., S. J. Tsai, and N. E. Huang. 2011. Decomposing the association of completed suicide with air pollution, weather, and unemployment data at different time scales. *Journal of Affective Disorders* 129 (1–3):275–81. doi: 10.1016/j.jad.2010.08.010.

Zhao, Y., Y. Wang, X. Zhang, and L. Wang. 2018. Exploring scale-specific controls on soil water content across a 500-Kilometer transect using multivariate empirical mode decomposition. *Vadose Zone Journal* 17 (1):180097. doi: 10.2136/vzj2018.05.0097.

Zou, H. 2006. The adaptive lasso and its oracle properties. *Journal of the American Statistical Association* 101 (476):1418–29. doi: 10.1198/016214506000000735.